Evaluating the New Keynesian Phillips Curve under VAR-based learning

Luca Fanelli

Department of Statistics, University of Bologna

January 2007

Evaluating the New Keynesian Phillips Curve under VAR-based learning*

Luca Fanelli†

University of Bologna

January 2007

Abstract

This paper proposes the evaluation of the New Keynesian Phillips Curve (NKPC) under a new learning mechanism where VAR learning dynamics is combined with the idea of testing the validity of the forward-looking model of inflation dynamics. The key assumption is that agents' perceived law of motion is a VAR whose parameters are updated by recursive least squares. Differently from standard adaptive learning methods, agents test sequentially the cross-equation restrictions that the NKPC imposes on the VAR as the information set increases. When the restrictions are not rejected agents learn under the restricted system and exploit the cross-equation restrictions to forecast inflation. It is thus possible to check how much and in which periods agents' beliefs are consistent with the restrictions of the theory. The empirical analysis on quarterly data on the euro area shows that the NKPC with negligible backward-looking parameter is not rejected when the model is evaluated over the period 1984-2005 under the proposed learning mechanism. The result, however, is not fully robust to specifications based on non stationary variables and points out that learning may represent a remarkable source of euro area inflation persistence but not its only determinant.

Keywords: Adaptive learning, Cross-equation restrictions, Forward-looking model of inflation dynamics, Perceived Law of Motion, Recursive Least Squares, VAR.

J.E.L. Classification: C32, C52, D83, E10

*Paper presented at the 17th EC^2 Meeting, "The Econometrics of Monetary Policy and Financial Decision-Making", December 15-16, 2006, Rotterdam, and at the 7th Macroeconometric Workshop: "Is macroeconometrics back?", Halle, December 7-8, 2006. I wish to thank Thomas Sargent, James Hamilton, Lutz Kilian, Kirstin Hubrich, Sophocles Mavroeidis, Barbara Rossi and Silvia Sgherri for helpful comments and suggestions on earlier drafts of the paper. I also thank Davide Delle Monache and Giulio Palomba for precious comments.

†Department of Statistical Sciences, University of Bologna, via Belle Arti 41, I-40126 Bologna, Italy. email: fanelli@stat.unibo.it, ph: +39 0541434303, fax: +39 051 232153.

1 Introduction

The class of 'small-scale' and 'medium-scale' dynamic stochastic general equilibrium (DSGE) models that presently govern the debate in monetary policy are built under the hypothesis of rational expectations (RE). Following Muth (1961), 'rational' agents compute expectations by referring to the 'true' probabilistic structure of the data. This means that expectations are formed with reference to the reduced-form model solution.

A controversial question is whether DSGE models are capable to capture observed data properties and persistence. Smets and Wouters (2003), among many others, argue that a DSGE model for the euro area exhibits fit and forecasting performance comparable to that of Vector Autoregressive (VAR) models, see also Ireland (2004). On the other hand, other authors recognize that DSGE models are potentially misspecified and propose methods to address the issue, e.g. Del Negro and Schorfheide (2006).

There is growing awareness among macroeconomists and econometricians, that the RE hypothesis requires too much knowledge. In practise, agents display 'bounded' rationality and depart systematically from Muth's tenet, see e.g. Pesaran (1987), Chap 3, Sargent (1999), and Evans and Honkapohja (1999, 2001), see also references therein. In the monetary policy framework the idea that inflation expectations may not be rational and that deviations from the RE hypothesis may represent a remarkable source of inflation persistence that may improve the fit of DSGE models is well recognized, see, *inter alia*, Roberts (1997) and Milani (2005*a*, 2005*b*).[1]

How to replace RE in econometric modelling, however, is still an open door. The traditional approach to modelling boundedly rational expectations assumes agents form expectations by using adaptive updating rules, see e.g. Branch and Evans (2006). In short, agents estimate and update the parameters of their forecasting model - the perceived law of motion (PLM) - according to recursive least squares (RLS), or a constant version of it, i.e. constant gain least squares (CGLS), see e.g. Sargent (1999).[2] Replacing RE in the forward-looking model by the forecasts implied by the PLM, yields the so-called Actual Law of Motion (ALM), which reads as the data generating process (DGP).

This paper focuses on the aggregate supply equation of a prototypical 'trinity' DSGE model of monetary policy, i.e. the New Keynesian Phillips curve (NKPC), and puts forward an econometric approach to the estimation and evaluation of the NKPC based on the combination of

[1] We also refer to e.g. to Sargent (1999), Orphanides and Williams (2005) and Primicieri (2006) for examples in the macroeconometric literature where rational policy-makers learn about the behavior of the economy in real time and set stabilization policies conditional on their current beliefs.

[2] In this context learning is 'adaptive' rather than 'optimal', because it ignores the feedback from the learning rule on the actual low of motion. In this paper we focus on the issue of learning from a purely econometric point of view.

the following assumptions: **(i)** agents use VARs to forecast inflation and the other variables of the system such as the labor share and the interest rate; **(ii)** agents evaluate the validity of the NKPC by testing the cross-equation restrictions (CER) that it imposes on the VAR for the data; **(iii)** agents update VAR parameters through RLS as new information become available and follow the outcome of a sequential test: if at time t the CER are rejected they forecast inflation from the unrestricted VAR; if at time t the CER are not rejected they forecast inflation from the constrained VAR, with the possibility of learning 'how much' and in which periods their statistical (forecast) model supports the restrictions of the theory.

The combination of **(ii)** and **(iii)** represents the major point of departure from existing literature. Models with adaptive learning are usually estimated by treating the ALM as the DGP, i.e. without any concern on the issue of testing whether the forward-looking model is consistent with data properties. More specifically, in the existing econometric literature once expectations are replaced by agents' beliefs, the ALM is cast in state space form and likelihood-based (possibly Bayesian) methods are applied to estimate the model, see e.g. Milani (2005*b*). The proposed learning mechanism, instead, takes an explicit stand on the idea that agents evaluate whether the ALM can be regarded as a statistically coherent description of the data.

We discuss in detail each of the three assumptions comprising our VAR-based model of learning dynamics.

Assumption (i): the hypothesis of 'VAR expectations' is not new in the literature, see Brayton et al. (1997) and Branch (2004). 'VAR expectations' can coincide or not with RE, depending on the actual model solution. We *do not assume* that the VAR, i.e. agents' PLM *necessarily coincides* with the Minimum State Variable (MSV) solution of the system (McCallum, 1983, 2003), as in our framework it reads as the 'best' statistical model agents specify to forecast the data, consistently with the information set they have.[3]

Assumption (ii): it is well known, at least since Sargent (1979), that models involving forward-looking behavior impose a set of CER on the VAR of the data. These restrictions can be used to identify and estimate the free parameters of structural forward-looking model, and to test their validity (Hansen and Sargent 1980, 1981). Further examples where VAR models are used as statistical platforms to test the restrictions implied forward-looking models include Baillie (1989), Johansen and Swensen (1999), Kozicki and Tinsley (1999), Fanelli (2002, 2006*a*, 2006*b, 2006c*) and Kollmann (2006).

Assumption (iii): VAR parameters are updated recursively as new data increase agents'

[3]This means that learning may be incomplete, see Sargent (1999). Observe that in this paper we are not interested in the conditions under which the economy converges to a RE equilibrium, see e.g. Marcet and Sargent (1992). We refer to Evans and Honkapohja (1999, 2001) and references therein for the study of convergence of different learning mechanisms, in a variety of macroeconomic models, to a unique equilibrium.

information set. Given the recursive nature of estimation and the assumption (ii) above, a sequence of CER arises. The objective of testing the sequence of CER can be accomplished by adapting Inoue and Rossi's (2005) sequential test. The asymptotic results in Inoue and Rossi (2005) cover Wald, Lagrange multiplier (LM) and likelihood ratio (LR)-like tests based on generalized method of moments (GMM). Therefore, included in this class are tests based on maximum likelihood that we consider in the paper. Learning in this set-up works as follows: if a time t the restrictions that the NKPC imposes on the VAR are not rejected, then it makes sense for the agents to forecast inflation by the constrained system, otherwise they will forecast from the unrestricted system.

Although the literature provides several examples where DSGE models comprising NKPC-type supply equations are investigated under the assumption of adaptive learning, to our knowledge, Milani (2005a) is the only contribution where 'the econometric analysis of the NKPC under learning rules' is explicitly addressed. Focusing on the US economy, Milani (2005a) considers specifications of the NKPC which account for the role of indexation, and a univariate dynamic forecast models which are estimated recursively by a constant gain version of least squares. He finds that when such a learning mechanism replaces the assumption of RE, a structural source of inflation persistence such as indexation is no longer essential to fit the data. Learning is thus interpreted as the major source of persistence in inflation, with consequences on the design of optimal monetary policy where the management of expectations plays a prominent role, see e.g. Woodford (2003) and Evans and Honkapohja (2003a. 2003b).

Our approach differs from Milani (2005a) in several respects. First, we use a VAR as the forecast model. Second, we update the VAR parameters through RLS, with the possibility of exploiting the link between RLS and maximum likelihood (ML) estimation. Third, and more importantly, our approach hinges on the idea of *testing* the NKPC and the learning mechanism is intimately related to the empirical assessment of the forward-looking model of inflation dynamics. Fourth, we extend the analysis to the case where the variables of the NKPC can be approximated as non stationary time-series, providing methods to check the sensitivity of results obtained by treating the PLM as a stationary system to specifications where possible stochastic trends characterizing variables are opportunely removed.

The method is applied to evaluate the NKPC in the euro area. We proxy firms' real marginal costs by the wage share as in Galí et al. (2001). The empirical analysis on quarterly data covering the 1971-2005 period shows that the NKPC is clearly rejected when a full sample ('one shot') test of the model is considered. Yet, the cross-equation restrictions with negligible backward-looking parameter are supported by the data when the VAR-based learning mechanism and the implied sequential test are taken into explicit account over the period 1984-2005, which is characterized by a marked disinflation process that facilitated the adoption of a single currency. The result, however, is not fully robust to specifications of the system based on non stationary variables and points out that: (a) there is little room for the NKPC to serve as a reasonable description of inflation dynamics under the rational expectations paradigm; (b) learning may represent a possible source of euro area inflation persistence but not its only determinant; (c) non stationarity can be a relevant issue.

The rest of the paper is organized as follows. Section 2 introduces the NKPC and Section 3 sketches the main features of the proposed method by discussing a motivating example. Section 4 proposes the econometric analysis of the NKPC. Section 4.1 deals with the derivation of CER between the VAR and the NKPC, whereas Section 4.1 extends the analysis to the case where agents behave as econometricians and update VAR parameters recursively and test the NKPC as the information set increases. Section 5 shows how the method can be extended to the case where variables are non stationary. Section 6 investigates the NKPC in the euro area; Section 6.1 deals with the empirical assessment of the model under VAR-based learning and Section 6.2 provides a robustness check of results. Section 6 contains some concluding remarks and the directions of future research. All proofs and technical details are sketched in the Appendix.

2 Model

A variety of pricing environments within the New Keynesian tradition give rise to the following 'hybrid' version of the NKPC (Roberts, 1995):

$$\pi_t = \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1} + \lambda x_t + u_t \tag{1}$$

where π_t is the inflation rate at time t, x_t is a scalar explanatory variable related to firms' real marginal costs, and $E_t\pi_{t+1}$ indicates the expected value of π_{t+1} formed at time t on the basis of the available information summarized in the (non decreasing) sigma-field $\Omega_t \subseteq \Omega_{t+1}$, i.e. $E_t\pi_{t+1} := E(\pi_{t+1} \mid \Omega_t)$. γ_f, γ_b and λ are the structural parameters. The model includes a disturbance u_t that is assumed to obey a martingale difference sequence (MDS) with respect to Ω_t.[4] Theory usually relates γ_f, γ_b and λ to other 'deep' parameters; for instance, in the Calvo model of Galí and Gertler (1999) and Galí et al. (2001), γ_f, γ_b and λ can be linked to firms' discount factor, the degree of price stickiness and the degree of 'backwardness' in price setting, and $\gamma_f + \gamma_b \leq 1$.

The empirical assessment of (1) has attracted a great deal of research, see, *inter alia*, Fuhrer and Moore (1995), Fuhrer (1997), Sbordone (2002, 2005), Galí et al. (2005), Lindé (2005), Rudd

[4] We explicitly rule out the assumption that u_t obeys some autoregressive (e.g. AR(1)) process, as it is customary in the literature.

and Whelan (2005*a*, 2005*b,* 2006), Kurmann (2006) and Fanelli (2006*c*), just to mention few. It is well understood in the recent literature that identification issues can be ruled out in the context of the forward-looking model (1) to the extent that a suitable dynamic system for x_t is explicitly specified, see e.g. Pesaran (1987), Chap. 6, Mavroeidis (2005) and Nason and Smith (2005). To achieve this objective one can in principle augment the model by introducing additional structural (possibly forward-looking) equations explaining the dynamics of x_t. This leads to the class of 'small scale' DSGE models of monetary policy, as in e.g. Lindè (2005). Alternatively, one can refer to reduced-form equations for x_t (and possibly for the other variables comprising the system), as in e.g. Fuhrer and Moore (1995) and Fuhrer (1997), who appeal to VAR-type dynamics.

Two likelihood-based approaches have been traditionally followed to investigate the NKPC. The former requires an explicit specification of the process generating x_t, and consists in the derivation of the reduced-form solution of the system under RE. The process generating x_t can be either a reduced-form as in e.g. Pesaran (1987), Chap 7, or a structural equation of a 'small scale' DSGE model, as in e.g. Lindè (2005). In both cases one can write down the log-likelihood of the system under the implicit null that the NKPC is the data generating process (DGP). In this framework it is of key importance to known whether the model solution is determinate or indeterminate, i.e. whether the system has a unique (non-explosive) solution, or multiple (non-explosive) solutions (Lubik and Schorfheide, 2004).

The latter approach, based on what we call 'VAR expectations' (Brayton et al., 1997) and that goes back to Sargent (1979), Campbell and Shiller (1987) and Baillie (1989), works under the assumption that irrespective of whether the NKPC (1) holds or not, the DGP for $Z_t := (\pi_t \colon x_t)'$ belongs to the VAR used by agents to forecast inflation and the other variables of the system, see e.g. Kurmann (2006) and Fanelli (2006*c*). Provided that the specified VAR reads as a congruent representation of the data, the idea is that if the theoretical model is 'true', its solution must be nested within agents' statistical model. The application of the method of undetermined coefficients allows to retrieve a set of CER between VAR and NKPC coefficients, that can be used to identify the structural parameters and for inferential purposes. Clearly, in practise there is no way to rule out that a class of possible solutions of the RE model might not be nested within the specified VAR, including e.g. the Minimum State Variable (MSV) solution (McCallum, 1983).[5]

In this paper we focus on the second approach and attempt to extend it to the case where the parameters of the VAR are updated recursively. Section 3 provides a motivating example

[5] Obviously, the two approaches are equivalent and end up with the same results when the two conditions are jointly met: first, the DGP is given by (1) (i.e. the NKPC holds); second, the implied model solution under RE is nested within the VAR used to approximate the dynamics of variables.

with the aim of illustrating the main features of the suggested method.

3 Background and outline of methodology

To fix main ideas, we leave for the moment the NKPC of Section 2 and focus on a more general example. Consider the forward-looking model

$$y_t = \gamma E_t y_{t+1} + \delta y_{t-1} + \kappa w_t + v_t \tag{2}$$

where y_t is a scalar, v_t a scalar white noise shock, and w_t a scalar explanatory variable; $E_t y_{t+1}$ indicates the expected value of y_{t+1} formed at time t on the basis of the available information, Ω_t, and γ, δ and κ are the structural parameters.

Many forward-looking models, including the NKPC (1), can be regarded as special cases of (2). To apply 'full' information methods to estimate (2) under RE, it is of key importance to know the structure of the process generating w_t, otherwise estimation through 'limited' information methods requires identification robust methods, see e.g. Mavroeidis (2006).

Suppose that boundedly rational agents behave as econometricians and have a PLM of the form

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 w_{t-1} + \eta_t \tag{3}$$

where η_t is a white noise process, and ϕ_0, ϕ_1 and ϕ_2 are unknown parameters. Agents gradually learn the values of ϕ_0, ϕ_1 and ϕ_2 from the data they observe. Notice that with $\phi_2 := 0$ in (3), agents would not recognize the dependence of y_t on w_{t-1}, hence in general agents, like econometricians, might fail to correctly specify the 'correct' model, and accordingly the ALM, see e.g. Evans and Honkapohja (1999), Section 5.3, for a comprehensive discussion.

At time t agents' one-step ahead forecasts (expectations) of y_t are formed by

$$\widehat{E}_t y_{t+1} := E(y_{t+1} \mid \mathcal{H}_t) := \widehat{\phi}_{0t} + \widehat{\phi}_{1t} y_t + \widehat{\phi}_{2t} w_t \tag{4}$$

where $\mathcal{H}_t := \{y_t, w_t, y_{t-1}, w_{t-1}, ..., y_1, w_1\} \subseteq \mathcal{H}_{t+1}$ is the information set they exploit at time t, and $\widehat{\phi}_{0t}$, $\widehat{\phi}_{1t}$ and $\widehat{\phi}_{2t}$ are obtained by estimating (3) by least squares, using $\mathcal{H}_t$. Plugging (4) back into equation (2) and solving for y_t, gives the ALM under least squares learning, i.e.[6]

$$y_t = (1 - \gamma\widehat{\phi}_{1t})^{-1}\widehat{\phi}_{0t}\gamma + (1 - \gamma\widehat{\phi}_{1t})^{-1}\delta y_{t-1} + (1 - \gamma\widehat{\phi}_{1t})^{-1}(\gamma\widehat{\phi}_{2t} + \kappa)w_t + v_t \tag{5}$$

where $\widehat{\varsigma}_t := (\widehat{\phi}_{0t} : \widehat{\phi}_{1t} : \widehat{\phi}_{2t})'$ evolves recursively with t.

[6] Observe that it is often assumed that when forming expectations in period t, in practise agents have access to information only up to $t-1$; in that case the quantity to plugg into (2) is not (4) but $\widehat{E}_{t-1}y_{t+1} := E(y_{t+1} \mid \mathcal{H}_{t-1})$. However, for the purposes of the present section, it is irrelevant which option among $\widehat{E}_{t-1}y_{t+1}$ and $\widehat{E}_t y_{t+1}$ is used in the analysis.

Two methods can be used to estimate a model of the form (5) whose coefficients are convolutions of the structural parameters γ, δ, κ and the time-varying coefficients representing agents' beliefs, $\widehat{\varsigma}_t := (\widehat{\phi}_{0t} : \widehat{\phi}_{1t} : \widehat{\phi}_{2t})'$. One possibility is to compute the one-step ahead forecasts of y_t by using (4), which are then plugged in back into (2); the resulting model can be estimated by standard methods, see e.g. Milani (2005a). The other possibility is to cast (5), along with the equations governing the recursion of $\widehat{\varsigma}_t$, in state space form and exploit likelihood-based (possibly Bayesian) techniques.

In line with the existing econometric literature on adaptive learning, the above mentioned estimation methods implicitly maintains that the forward-looking model (2) holds: once RE in (2) are replaced by agents' beliefs, the only problem is the estimation of the resulting time-varying ALM. No attention is devoted on the issue of testing whether the structural model (2) represents a plausible description of the data under the chosen PLM and updating rule.

The model we propose in this paper is based on the idea that at each time t, agents acting like econometricians use their PLM to test whether the restrictions implied by the forward-looking model are supported by the data. The implicit assumption, therefore, is that the chosen PLM represent a congruent description of the data and the DGP belongs to it. If the restrictions match the data, it makes sense to use the ALM to forecast the variable(s) of interest, otherwise agents' 'best' forecast will be that stemming from the PLM. This procedure can be iterated at time $t+1$, $t+2$,... with the increase of the information set, so that other than updating their beliefs, agents can also learn to what extent and in which periods their PLM embodies information on the theoretical model.

To describe how such a method may work in practise with reference to the model (2), suppose that boundedly rational agents have a PLM for the bivariate vector $Z_t := (y_t \colon w_t)'$ which is a VAR(k), where k, the number of lags, is determined by the data. Following Sargent (1979), Campbell and Shiller (1987) and Baillie (1989), the VAR can be regarded in this framework not only as the forecast system, but also as the statistical model for the data upon which the restrictions implied by the theory can be tested. Thus if at time t the CER that the model (2) imposes on the VAR(k) are not rejected, then it means that agents' PLM supports the restrictions of the theory, and it makes sense to forecast y_t by the restricted VAR(k) (which amounts to use the ALM). Likewise, if at time t the CER are rejected, it is reasonable to forecast y_t by the unrestricted VAR(k), and wait for new data at time $t+1$ to repeat the test of the model, and so forth. In other words, we are describing a learning process based on a sequential test of the validity of the CER that the theoretical model imposes on the VAR.

Consider, in particular, the VAR(2)

$$Z_t = G_1 Z_{t-1} + G_2 Z_{t-2} + \varepsilon_t \tag{6}$$

where ε_t is a bivariate white noise and G_i, $i = 1, 2$, are 2×2 matrices of parameters. VAR forecasts are given by $\widehat{E}_{t-1}y_{t+1} := E(y_{t+1} \mid \mathcal{H}_{t-1}) = g_y G_t^2 \widetilde{Z}_{t-1}$, where with G_t we conventionally denote the companion matrix associated with (6) to be estimated by using the information up to time t, $\mathcal{H}_t := \{Z_t, Z_{t-1}, ..., Z_1\} \subseteq \Omega_t$, $\widetilde{Z}_t := (Z_t' \vdots\, Z_{t-1}')'$, and g_y is a selection row vector such that $g_y \widetilde{Z}_t := y_t$. The forward-looking model (2) can be tested at time t as follows: (i) condition both sides of (2) with respect to the information set at time $t-1$ and apply the law of iterated expectations, obtaining

$$E_{t-1}y_t = \gamma E_{t-1}y_{t+1} + \delta y_{t-1} + \kappa E_{t-1}w_t;$$

(ii) substitute $E_{t-1}y_t$, $E_{t-1}y_{t+1}$ and $E_{t-1}w_t$ in the expression above by the VAR(2) forecasts $\widehat{E}_{t-1}y_t$, $\widehat{E}_{t-1}y_{t+1}$ and $\widehat{E}_{t-1}w_t$, yielding

$$g_y G_t \widetilde{Z}_{t-1} = \gamma g_y G_t^2 \widetilde{Z}_{t-1} + \delta g_y Z_{t-1} + \kappa g_w G_t \widetilde{Z}_{t-1}$$

where g_w is a known selection vector such that $g_w Z_t := w_t$. Since $\widetilde{Z}_{t-1} \neq 0$ a.s., one has

$$g_y G_t (I_4 - \gamma G_t) - \delta g_y - \kappa g_w G_t = \underset{1\times 2}{0} \tag{7}$$

which involve both the structural parameters of (2), and the VAR parameters. Thus for $t = T_0 + 1, T_0 + 2, ...$ we have the sequence of CER

$$g_y G_t (I_4 - \gamma G_t) - \delta g_y - \kappa g_w G_t = \underset{1\times 2}{0} \quad , \quad t = T_0 + 1, T_0 + 2, ... \tag{8}$$

where $T_0 + 1$ can be regarded as the 'first monitoring time', i.e. the first period in which the evaluation of the CER starts.

Provided that the VAR is identified under the CER, a natural solution to test (8) is to estimate, for $t = T_0 + 1, T_0 + 2, ...$, the system (6) both unrestrictedly and under the CER and hence compute a sequence of LR statistics. In the next sections we show in detail how the method works in practise with the NKPC.

4 Expectations formation model and the cross-equation restrictions

In this section we introduce the expectations formation model and discuss the CER that the NKPC imposes on the VAR parameters. To keep the exposition simple, we divide the presentation into two parts. In Section 4.1 we derive the CER assuming that all the available information, $\mathcal{H}_{T^{\max}}$, where $T^{\max}$ denotes the time of the last available information, is used to estimate and test the model in a 'one-shot' solution. In Section 4.2 we extend the analysis to the case where agents endowed with initial VAR estimates obtained with data up to time T_0 update parameter values recursively and test the restrictions implied by the NKPC at time $t = T_0 + 1, T_0 + 2$ ...., $T^{\max}$.

4.1 VAR approach without learning

Assume that agents' law of motion for the $p \times 1$ vector of observable variables, Z_t, is given by

$$Z_t := \sum_{i=1}^{k} A_i Z_{t-i} + \Theta d_t + \varepsilon_t \tag{9}$$

where k is the lag length, Z_0, $Z_{-1}, ..., Z_{(1-k)}$ are fixed, A_i, $i = 1, 2, ..., k$ are $p \times p$ matrices of parameters, d_t is an $d_0 \times 1$ vector of deterministic terms (constant, linear trend, deterministic dummies, etc.) with associated $p \times d_0$ matrix of parameters, Θ, and ε_t is a MDS with respect to $\mathcal{H}_t := \sigma(Z_t, Z_{t-1}, ..., Z_1) \subseteq \Omega_t$, with (non-singular) covariance matrix Σ_ε and Gaussian distribution.

If the VAR is (asymptotically) stable, i.e. the roots of $\det(I_p - \sum_{i=1}^{k} A_i s^i) = 0$ are such that $\mid s \mid > 1$, ℓ-step ahead forecasts of Z_t can be computed as

$$\widehat{E}_t Z_{t+\ell} := E(Z_{t+\ell} \mid \mathcal{H}_t) := g_z A^\ell \widetilde{Z}_t + \sum_{h=0}^{\ell-1} g_z A^h g_z' \Theta d_{t+\ell-h} \tag{10}$$

where $\widetilde{Z}_t := (Z_t' \colon Z_{t-1}' \colon ... \colon Z_{t-k+1}')'$ is the $n \times 1$ $(n = pk)$ state vector associated with the VAR (9), A is the $n \times n$ companion matrix, and g_z is a $p \times n$ selection matrix such that $g_z \widetilde{Z}_t := Z_t$.

Suppose, in particular, that $Z_t := (\pi_t \colon x_t \colon a_t')'$, where π_t is inflation, x_t a proxy of firms' real marginal costs, and a_t is a $q_a \times 1$ $(p = 2 + q_a)$ subvector of 'additional' variables that possibly help to forecast x_t or that play an important role in the system; in Section 6 a_t will be a scalar $(q_a := 1)$ representing a short term interest rate.

From (10) it turns out that with $\ell := 1$ and $\ell := 0$

$$\widehat{E}_{t-1}\pi_{t+1} \ : \ = E(\pi_{t+1} \mid \mathcal{H}_{t-1}) := g_\pi A^2 \widetilde{Z}_{t-1} + g_\pi \Theta d_{t+1} + g_\pi A g_\pi' \Theta d_t \tag{11}$$

$$\widehat{E}_{t-1}\pi_t \ : \ = E(\pi_t \mid \mathcal{H}_{t-1}) := g_\pi A \widetilde{Z}_{t-1} + g_\pi \Theta d_t \tag{12}$$

$$\widehat{E}_{t-1}x_t \ : \ = E(x_t \mid \mathcal{H}_{t-1}) := g_x A \widetilde{Z}_{t-1} + g_x \Theta d_t \tag{13}$$

where g_π and g_x are $1 \times n$ selection vectors such that $g_\pi \widetilde{Z}_t := \pi_t$, and $g_x \widetilde{Z}_t := x_t$. Condition both sides of the NKPC (1) with respect to $\mathcal{H}_{t-1}$, and apply the law of iterated expectations, obtaining

$$E(\pi_t \mid \mathcal{H}_{t-1}) = \gamma_f E(\pi_{t+1} \mid \mathcal{H}_{t-1}) + \gamma_b \pi_{t-1} + \lambda E(x_t \mid \mathcal{H}_{t-1}); \tag{14}$$

substituting the VAR forecasts (11)-(13) into (14) and ignoring deterministic components in d_t,[7] yields the following set of restrictions

$$g_\pi A(I_n - \gamma_f A) - \gamma_b g_\pi - \lambda g_x A = 0 \tag{15}$$

involving the structural parameters of the forward-looking model and the VAR parameters A_i, $i = 1, 2, ..., k$.

Proposition 1 of Appendix A[8] discusses the technical conditions under which there exists a unique explicit form representation of the constraints (15) which ensure the (local) identifiability of the restricted VAR. In particular, it is shown that if the 'true' value of λ is different from zero, it is possible to express the parameters associated with the VAR marginal model for x_t, B_x, as unique function of the structural parameters, $\tau = (\gamma_f : \gamma_b : \lambda)'$, and the other VAR parameters; formally:

$$B_x := f(B_\pi, B_a, \tau) \tag{16}$$

where f is a non-linear vector function, B_π and B_a are the vectors containing the VAR parameters associated with the equations for π_t and a_t, respectively (except the coefficients associated to deterministic components), and the non-linear relation (16) holds in a neighbourhood of the 'true' parameter values.

The maximization of the VAR log-likelihood subject to (16) can be carried out by numerical optimization algorithms. Fanelli (2006*c*) proposes a combination of a grid-search for the structural parameters τ with 'standard' (quasi) Newton-type algorithms, whereas Kurmann (2006) recommends the simulated annealing algorithm (e.g. Goffe et al., 1994). In both cases LR tests for the CER can be computed.

4.2 VAR approach with learning

In this section we extend the approach discussed in Section 4.1 one step further by recognizing that in computing the forecasts (10) at time t, in practise agents use only the information set $\mathcal{H}_t$. This fact has direct consequences on the way the NKPC is tested.

We split the sample into two parts: the first, a pre-testing period from $t = 1$ to $t = T_0$, is the period used to produce initial estimates of VAR parameters; the second part, from $t = T_0 + 1$ onward is the sample used to evaluate the NKPC through the CER. Observe that T_0+1 must be

[7] We ignore deterministic components for simplicity, hence the derivation of CER can be interpreted with reference to a PLM whose variables are expressed in terms of deviations from determinist terms. In general, it is possible to take into account the role of deterministic terms provided that d_t is opportunely included also in the NKPC (1), otherwise the derivation of CER becomes burdensome.

[8] It generalizes Proposition 2 in Kurmann (2006) and is based on a completely different proof method.

'reasonably' large in the sense that the estimates of the VAR parameter $(A_1 \colon A_2 \colon \ldots \colon A_k \colon \Theta : \Sigma_\varepsilon)$ based on $\mathcal{H}_{T_0+1}$ must ensure that a second-order moment matrix of parameters, defined below, is invertible. We further assume that T_0+1 is large enough to guarantee that the estimates of $(A_1 \colon A_2 \colon \ldots \colon A_k \colon \Theta : \Sigma_\varepsilon)$ based on least squares are consistent; we refer to Carceles-Poveda and Giannitsarou (2006) for a general discussion on the effect of initial conditions on adaptive learning algorithms.

If at time $t \geq T_0 + 1$ agents behave as econometricians, their actual forecasts will be

$$E(Z_{t+\ell} \mid \mathcal{H}_t) := g_z' \ \widehat{A}_t^\ell \widetilde{Z}_t + \sum_{h=0}^{\ell-1} g_z \widehat{A}_t^h g_z' \widehat{\Theta}_t d_{t+\ell-h} \ , \qquad \begin{array}{c} \ell = 1, 2, \ldots \\ t = T_0 + 1, T_0 + 2 \ldots \end{array} \tag{17}$$

where $\widehat{A}_t$ and $\widehat{\Theta}_t$ are the estimators of A and Θ obtained from $\mathcal{H}_t$.

Let $B := (A_1 \colon A_2 \colon \ldots \colon A_k : \Theta)$ be the $p \times u$, $u = n + d_0$, matrix of first-moment parameters associated with the VAR (9). Focusing on the case $\ell := 1$ (one-step ahead predictions), (17) specializes in

$$E(Z_{t+1} \mid \mathcal{H}_t) = Z_{t+1|t}^f = \widehat{B}_t \ X_{t+1}^*$$

where $X_t^* := (Z_{t-1}' \colon Z_{t-2}' \colon \ \ldots \colon Z_{t-k}' : d_t)'$ is $u \times 1$, and given the data up to time t, the LS estimator of B and Σ_ε are given, respectively, by

$$\widehat{B}_t := \left(\frac{1}{t} \sum_{i=1}^{t} Z_i X_i^{*\prime} \right) \left(\frac{1}{t} \sum_{i=1}^{t} X_i^* X_i^{*\prime} \right)^{-1} := M_t^{zx^*} (M_t^{x^*x^*})^{-1} \tag{18}$$

$$\widehat{\Sigma}_{\varepsilon t} := \frac{1}{t} \sum_{i=1}^{t} (Z_i - \widehat{B}_t X_i^*)(Z_i - \widehat{B}_t X_i^*)'. \tag{19}$$

It is worth noting that under the assumption of Gaussian ε_t, the expressions in (18) and (19) correspond to the maximum likelihood (ML) estimates of B and Σ_ε obtained through $\mathcal{H}_t$. Moreover, the value of the log-likelihood at the maximum is given, a part from a constant, by

$$\log \widehat{L}_t^{\max} = -\frac{t}{2} \log(\det(\widehat{\Sigma}_{\varepsilon t})) \quad , \quad t = T_0 + 1, T_0 + 2 \ldots \tag{20}$$

Rewrite the expression (18) in the form

$$\widehat{B}_t M_t^{x^*x^*} := M_t^{zx^*}, \quad t = T_0 + 1, T_0 + 2 \ldots .$$

Using the recursive relations

$$\begin{aligned} M_t^{x^*x^*} \quad &: \quad = \frac{t-1}{t} M_{t-1}^{x^*x^*} + \frac{1}{t} X_t^* X_t^{*\prime} \\ M_t^{zx^*} \quad &: \quad = \frac{t-1}{t} M_{t-1}^{zx^*} + \frac{1}{t} Z_t X_t^{*\prime} \end{aligned}$$

and some routine algebra one gets[9]

$$\widehat{B}'_t \quad : \quad = \widehat{B}'_{t-1} + \frac{1}{t}(M_t^{x^*x^*})^{-1}X_t^*(Z_t - \widehat{B}_{t-1}X_t^*)' \tag{21}$$

$$M_t^{x^*x^*} \quad : \quad = M_{t-1}^{x^*x^*} + \frac{1}{t}(X_t^*X_t^{*\prime} - M_{t-1}^{x^*x^*}) \tag{22}$$

$$t \quad = \quad T_0 + 1, T_0 + 2, ... \tag{23}$$

For given initial values of $\widehat{B}_{T_0}$ and $M_{T_0}^{x^*x^*}$, (21)-(23) define the Recursive LS (RLS) estimator of B.

Consider now the following partition of B, $B = [B_1 \colon B_2 \colon \ldots \colon B_p]'$, where each B_j, $j = 1, 2, ..., p$, is the $u \times 1$ vector containing the parameters of the j-th equation of the VAR. For instance, referring to the trivariate VAR $Z_t := (\pi_t \colon x_t \colon a_t)'$ already discusses in Section 4.1, one has $B = [B_1 \colon B_2 \colon B_3]'$, where $B_1' := [B_\pi' : \Theta_\pi']$, $B_2' := [B_x' : \Theta_x']$ and $B_3' := [B_a' : \Theta_a']$, and Θ_v', $v = \pi, x, a$ are the coefficients associated with deterministic terms in the three equations. The RLS (21) specializes in the well known relation

$$\widehat{B}_{jt} := \widehat{B}_{jt-1} + \frac{1}{t}(M_t^{x^*x^*})^{-1}X_t^*(Z_{jt} - \widehat{B}'_{jt-1}X_t^*) \quad , \quad \begin{array}{c} j = 1, 2, ..., p \\ t = T_0 + 1, T_0 + 2, ... \end{array} \tag{24}$$

where $\widehat{e}_{jt} := (Z_{jt} - \widehat{B}'_{jt-1}X_t^*) = (Z_{jt} - Z_{jt|t-1}^f)$ is the observed prediction error relative to the j-th variable (equation) of the VAR. The RLS updating of the second order moments matrix M_t^{xx} described in (22) is common to the p equations of the VAR.

In the literature on adaptive learning the recursion (24) if often replaced by the more general scheme

$$\widehat{B}_{jt} \quad : \quad = \widehat{B}_{jt-1} + g_{jt}(M_t^{x^*x^*})^{-1}X_t^*(Z_{jt} - \widehat{B}'_{jt-1}X_t^*) \tag{25}$$

$$M_t^{x^*x^*} \quad : \quad = M_{t-1}^{x^*x^*} + g_{jt}(X_t^*X_t^{*\prime} - M_{t-1}^{x^*x^*}) \ , \quad \begin{array}{c} j = 1, 2, ..., p \\ t = T_0 + 1, T_0 + 2, ... \end{array} \tag{26}$$

where each g_{jt} is a scalar, known as the 'gain' sequence. In general, $g_{jt} \neq g_{it}$ for $i \neq j$, i.e. the gains may differ across VAR equations. With $g_{jt} := t^{-1}$ the expressions (25)-(26) collapse to RLS (21)-(23); the position $g_{jt} := g_j$, $0 < g_j < 1$, leads to the constant gain version of recursive least squares (CGLS), where the term 'constant' stresses the difference with the previous case characterized by decreasing gains. For $j = 1, 2, ..., p$, CGLS discounts past observations at geometric rate $1 - g_j$, and is therefore more robust to structural changes, see e.g. Branch and

[9] In particular, for given matrices A, B, and C of suitable dimensions, we use the following matrix inversion rule: $(C + B'A^{-1}B)^{-1} := C^{-1} - C^{-1}B'(A + BC^{-1}B')^{-1}BC^{-1}$.

Evans (2006).[10] Given the 'natural' link between RLS and the recursive application of ML (provided that VAR disturbances are Gaussian), in this paper we focus on (21)-(23).[11]

The restrictions that the NKPC imposes on the VAR parameters can be derived by following the same route of Section 4.1. In this case one gets the sequence of CER:

$$g_\pi A_t(I_n - \gamma_f A_t) - \gamma_b g_\pi - \lambda g_x A_t = 0 \quad , \quad t = T_0 + 1, T_0 + 2, ... \tag{27}$$

where once again the time index attached to the companion matrix stresses that the estimation of the parameters in A is based on the information available up to time t, $\mathcal{H}_t$. By applying Proposition 1 of Appendix A, from (27) we can now derive the following sequence of explicit form constraints

$$B_{xt} := f(B_{\pi t}, B_{at}, \tau) \quad , \quad t = T_0 + 1, T_0 + 2, ... \tag{28}$$

which read as the 'sequential' counterpart of (16). Thus, under the CER the recursion of the estimators of $B_{\pi t}$ and B_{at} is governed by (24) and simultaneously the recursion of B_{xt} follows from $\widehat{B}_{xt} := f(\widehat{B}_{\pi t}, \widehat{B}_{at}, \tau)$, given the form of $f(\cdot)$.

The value of the likelihood at the constrained maximum is given (a part from a constant) by

$$\log \widetilde{L}_t^{\max} = -\frac{t}{2} \log(\det(\widetilde{\Sigma}_{\varepsilon t})) \quad , \quad t = T_0, T_0 + 1, ...$$

with $\widetilde{\Sigma}_{\varepsilon t}$ being the estimated covariance matrix of the restricted VAR at time t, thus the 'natural' likelihood-oriented approach to test the restrictions (28) consists in computing the LR statistics

$$LR_t = -2(\log \widetilde{L}_t^{\max} - \log \widehat{L}_t^{\max}) \quad , \quad t = T_0 + 1, T_0 + 2, ... \tag{29}$$

Apparently, to decide about the null that the CER hold, one might simply compare the LR statistics in (29) with the conventional χ^2-based large-sample critical values. Focusing on the GMM framework, however, Inoue and Rossi (2005) recently show that using conventional critical values, by the law of iterated logarithms the probability that such type of tests eventually leads to a rejection of the null hypothesis is asymptotically 1. To remedy this problem, they derive asymptotic critical values that allow one to 'follow' the test statistic through the whole sequence $t = T_0 + 1, T_0 + 2...$, in such a way that the probability of rejecting the null hypothesis is under control at each t. Inoue and Rossi (2005) also provide asymptotic critical values (their Table 1)

[10] In fact, compared to RLS constant gain algorithms assign decreasing weights on observations across time, so that recent observations matter for the current estimate even in the limit. For this reason it is commonly argued that CGLS are useful in tracking structural changes in the model, hence they are particular suited in situations where the policy is allowed to shift over time.

[11] We leave the extension of the proposed learning method to the case of CGLS to future research.

for LR-type sequential tests that can be used in the empirical analysis. The results in Inoue and Rossi (2005) cover Wald, Lagrange multiplier (LM) and likelihood ratio (LR)-like tests based on GMM, hence potentially included in this class is (29).

5 Non stationary variables

While theories upon which DSGE models and in particular the NKPC are built are formulated at the single agent level, estimation is usually based on aggregate data. Aggregation may have both theoretical and empirical consequences. For instance, as shown in e.g. Hughes Hallet (2000), the aggregation of sectoral, regional/national Phillips curves may yield an inflation-unemployment trade-off that is not vertical in the long run, despite the 'individual' curves being vertical. Likewise, aggregation may be a potential source of non stationarity; that this can be the case for the euro area inflation is confirmed by e.g. Benigno and López-Salido (2006) where the existence of heterogeneity in inflation persistence across regions is documented, and by O'Reilly and Whelan (2005) and Batini (2006), who using different techniques find that the persistence of inflation is close to one and stable over time.

It is well recognized that if the eigenvalues of the VAR companion matrix, A, are close to the unit circle, test statistics based on standard asymptotic theory and the typical sample lengths of macroeconomic analysis may suffer large size distortion and power losses, see e.g. Johansen (2006). In these circumstances, not fixing the number of unit roots of the system may worsen the small sample distortion. Thus, when there exists a sound suspect that the variables entering the NKPC (1) are non stationary and can be reasonably approximated in terms of processes integrated of order one (I(1)), a solution may be that of checking the sensitivity of results to specifications where the I(1) trends are opportunely removed.

When the variables entering agents' PLM are I(1), it is either possible that π_t and x_t are not cointegrated, or that they are cointegrated, see Fanelli (2006*c*) for a comprehensive discussion. In general, by allowing the roots of the characteristic equation $\det(I_p - \sum_{i=1}^{k} A_i s^i) = 0$ of the VAR (9) to be also at $s = 1$, one can disentangle empirically between the above mentioned situations by testing the cointegration rank (i.e. the number of unit roots) of the VAR, see e.g. Johansen (1996).

If π_t and x_t are I(1) and not cointegrated, two empirical formulations of the NKPC are consistent with data properties. First, it may happen that e.g. π_t is I(1) and x_t is I(0), that means that it exists a trivial cointegrating relation of the form $(0 : 1)Z_t$.[12] Such a situation can

[12] In principle it might also happen that π_t is I(0) and x_t is I(1), a situation which can be faced empirically by replacing x_t in (1) by Δx_t, and considering a VAR for $Z_t := (\pi_t : \Delta x_t : a_t)'$ (provided that also a_t is I(0)).

be accounted by manipulating (1) in the form

$$\Delta\pi_t = \frac{1-\gamma_b}{\gamma_b}E_t\Delta\pi_{t+1} + \frac{\lambda}{\gamma_b}x_t + u_t^* \tag{30}$$

which is obtained by simple algebra and imposing the restriction $\gamma_f + \gamma_b = 1$, with $u_t^* := \gamma_b^{-1}u_t$. The econometric evaluation of the model (30) under VAR-based learning follows the lines described in Section 4.2 for stationary models, the only exception being that now agents' PLM reads as a VAR(k) for $Z_t := (\Delta\pi_t : x_t : a_t)'$ (we have assumed that also a_t is I(0)).

Second, it may happen that both π_t and x_t are I(1) but do not share a common stochastic trend (including that of a_t). In this case the empirical analysis of the NKPC under VAR-based learning follows the route described in Section 4.2 for the case of stationary variables, provided that the procedure is opportunely adapted. In particular, the non stationary VAR(k) for $Z_t := (\pi_t : x_t : a_t)'$ must be replaced by its differenced counterpart, i.e. a DVAR($k-1$) for $\Delta Z_t := (\Delta\pi_t : \Delta x_t : \Delta a_t)'$ and at the same time the NKPC (1) must be empirically replaced by the 'accelerationist' specification

$$\Delta\pi_t = \gamma_f E_t\Delta\pi_{t+1} + \gamma_b\Delta\pi_{t-1} + \lambda\Delta x_t + u_t. \tag{31}$$

Differently from what happens with (30) and (32) (see below), the empirical formulation (31) of the NKPC can not be obtained from (1) by algebraic manipulations but by simply replacing the levels of the variables, expect u_t, with their first differences.

When π_t and x_t are I(1) and cointegrated, it is convenient to express the NKPC in a form such that both levels and first differences are involved. By simple algebra (1) can be reparameterized in the form

$$\Delta\pi_t = \omega E_t\Delta\pi_{t+1} - \psi(\pi_t - \xi\ x_t) + u_t^* \tag{32}$$

where

$$\xi \ := \frac{1}{(1-\gamma_f-\gamma_b)}\lambda \tag{33}$$

$$\omega \ := \frac{\gamma_f}{\gamma_b} \tag{34}$$

$$\psi \ := \left(\frac{1-\gamma_f-\gamma_b}{\gamma_b}\right) \tag{35}$$

$$u_t^* \ := \omega u_t$$

provided that $\gamma_f + \gamma_b < 1$. The main issue with (32), however, is the interpretation of the stationary process $(\pi_t - \xi\ x_t)$, given that theory does not predict that e.g. inflation and the wage share (unit labor costs) must be cointegrated for the NKPC to hold. Notice that in this set-up the structural parameter λ belongs to the cointegration space.

We refer to Fanelli (2006c) for a detailed discussion of how the econometric analysis of (32) can be developed under 'VAR expectations'. In short, testable implications of (32) can be derived by appealing to a stationary vector autoregressive representation of the vector $W_t := (\Delta\pi_t : e_t : \Delta a_t)'$, where $e_t := (\pi_t - \widehat{\xi}\ x_t)$, and $\widehat{\xi}$ is a super-consistent estimator of ξ in (33). The extension to the case of learning along the lines described in Section 4.2 is straightforward to the extent that agents (i) have some a priori knowledge on the fact that π_t and x_t are cointegrated, and possess some 'external' information about the value assumed by ξ in (33).

6 Results

We consider quarterly data relative to the euro area, using the last release of the Area-wide Model (AWM) data set described in Fagan et al. (2001). Variables cover the period 1971:1-2005:4. To measure inflation we use the GDP deflator, i.e. $\pi_t := 100(p_t - p_{t-4})$, where p_t is the log of the GDP deflator. As in Gali et al. (2001), firms' average marginal costs are proxied by the wage share (log of real unit labour costs), $x_t := ws_t$. A short term interest rate, $a_t := i_t$ ($q_a = 1$), is included in the system.[13]

The plot of the three variables is reported in Figure 1; the vertical line at 1984:1 indicates the beginning of the sub-sample used to evaluate the NKPC, whereas the period 1971:1-1983:4 is used to produce initial estimates of the VAR parameters. In terms of the notation of the previous sections, $T_0 + 1$:=1984:1 is the first monitoring period and $T^{\max}$:=2005:4 is the last available observation.

Note that the sample selected to produce initial estimates of the VAR parameters is characterized by an high inflation regime. As argued in Batini (2006), most European central banks adopted a more aggressive approach toward fighting inflation around the first half of the eighties. The sample used to evaluate the NKPC can be regarded as substantially homogeneous from the point of view of the commitment of European monetary authorities to stabilize inflation, the convergence towards the European Monetary Union, and the maintenance of a stable currency.

The simple graphical inspection of Figure 1 reveals, in line with O'Relly and Whelan (2005) and Batini (2006), that (aggregate) euro area inflation persistence resembles that of unit root-type processes.[14]Also the persistence of the wage share and of the short term interest rate appears

[13]The GDP deflator is YED in the AWM data set. The wage share (real unit labour costs) is computes as $ws_t := 100 \times \log(WIN_t/YER_t)$, where WIN is 'Compensation to Employees' (in real terms) and YER is real GDP. As a proxy of i_t we have used the short-term interest rate (nominal in percent), which is STN in the data set.

[14]Both O'Reilly and Whelan (2005) and Batini (2006) conclude that euro area inflation persistence is high, however the former finds a possible shift in inflation persistence around 1984, whereas the latter do not detect

considerably high: in general, such an outcome is not surprising in view of the cross-country aggregation process underlying data construction.

A VAR for $Z_t := (\pi_t\colon ws_t\colon i_t)'$ with $k := 3$ lags and a constant is estimated over the period 1971:1-1983:4. Table 1 reports some residuals diagnostic statistics, along with the highest estimated eigenvalues of the VAR companion matrix, and Johansen's (1996) LR trace test for cointegration rank.[15] It can be noticed that with $k := 3$, VAR residuals appear substantially 'well-behaved' over the sample 1971:1-1983:4, suggesting that RLS estimates can be reasonably approximated by the reiterated application of ML. Yet, results in Table 1 cast doubts on the stationary of the system. For this reason, after discussing the empirical evaluation of the NKPC in Section 6.1, we support the evaluation of the model through a robustness check in Section 6.2.

6.1 Euro area NKPC with learning dynamics

Given the initial estimates discussed above, the sequential evaluation of the NKPC under VAR dynamics outlined in Section 4.2 works by estimating the VAR(3) recursively over the period 1984:1-2005:4, both unrestrictedly and subject to the CER restrictions (28).

The maximization of the Gaussian likelihood under the CER is performed by calibrating the structural parameters $\tau = (\gamma_f\colon \gamma_b\colon \lambda)'$ at $\tau := \tau^* = (0.90\colon 0.05\colon 0.20)'$, which roughly correspond, in terms of the Calvo model of inflation dynamics, to firms' discount factor equal to 0.95, a fraction of 3% of backward-looking firms, and an average time of about 2.8 quarters over which prices are kept fixed. The idea of fixing τ to the value τ^* has a twofold motivation. First, we can not implement a grid search estimation for τ as in that case the asymptotic critical values of the sequential test of Inoue and Rossi (2005) would not apply. Second, fixing τ to τ^* has its ratio in the similar values of structural parameters $\tau^{**} = (0.773\colon 0.043\colon 0.214)'$ obtained by Gali et al. (2001) through GMM estimation and the AWM data set until 1998; in terms of the Calvo model τ^{**} corresponds, a part from the discount rate, to a fraction of 3% of backward-looking firms, and an average time of 3 quarters of sticky prices duration; see also Benigno and López-Salido (2006).

Figure 2 plots the LR statistic (LR1) for the CER over the period 1984:1-2005:4, along with both the 5% critical values computed following the method of Inoue and Rossi (2005), and the 0.95 quantile of a χ^2-distribution with 9 degree of freedom.[16] For comparative purposes Figure

any changes in persistence. Moreover, it is woth noting that Benigno and López-Salido (2006) find supporting evidence of the existence of heterogeneity in iflation dynamics across euro area countries.

[15] Calculations were performed in PcGive 10.0, Doornik and Hendry (2001).

[16] Observe that there are $pk := 9$ and not $pk - 3 := 6$ CER (see Proposition 1, Appendix A) since in this application the three structural parameters $\tau = (\gamma_f\colon \gamma_b\colon \lambda)'$ have been fixed.

2 also plots the LR statistic (LR2) for the CER obtained by fixing the structural parameters at τ^{**} as in Gali et al. (2001).

The graph reveals that if one focuses on the full sample ('one shot') test of the CER, i.e. on the value of LR1 at 2005:4, and compare it with the $\chi^2_{0.95}(9)$ quantile, the NKPC is sharply rejected. This result confirms e.g. the empirical findings in Fanelli (2006*c*), where using the AWM data set with variables available until 1998:2 and a VAR-based approach, the NKPC is sharply rejected. However, if one takes into explicitly account the sequential nature of the LR1 statistic and uses the 5% asymptotic critical values from Table1 in Inoue and Rossi (2005), it turns out that with $\tau := \tau^* = (0.90\colon 0.05\colon 0.20)'$ the NKPC is not rejected over the entire period 1984:1-2005:4; the NKPC would be rejected if τ were fixed at $\tau^{**} = (0.773\colon 0.043\colon 0.214)'$ (LR2 statistic) as in Gali et al. (2001).

Empirical results show that if agents form their beliefs recursively through the VAR(3) over the period 1984:1-2005:4, then their PLM supports the restrictions implied by the NKPC 100% of times. Following Milani (2005*a*), this finding can be interpreted by arguing that learning is the main source of euro area inflation persistence, given that the size of γ_b in τ^* is negligible compared to that of γ_f. On the contrary, within the RE paradigm it is usually found that further lags of inflation, reflecting indexation, rules of thumb, contracts, etc., are typically needed to match the data, see e.g. Fuhrer and Moore (1995) and Fuhrer (1997). We refer to e.g. Bullard and Mitra (2002), Woodford (2003) and Evans and Honkapohja (2003*a*. 2003*b*) and references therein for a detailed investigation of the optimal design of monetary policy rules when instead of having RE private agents follow plausible learning rules.

Observe, further, that the square root of the mean square forecast error of inflation, computed as $MSE(\pi) := (1/T_1)\sum_{t=T_0+1}^{T^{\max}}(\pi_t - \widehat{\pi}_{t|t-1})^2$, where $\widehat{\pi}_{t|t-1}$ is the one-step ahead forecast of inflation based on the information up to time $t-1$, and T_1 is the number of quarters between $T_0 + 1 = 1984 : 1$ and $T^{\max} = 2005 : 4$, is equal to 0.93 using the unrestricted VAR(3), and is equal to 0.98 using the VAR(3) under the CER, emphasizing that the forecast success of the PLM and the ALM is substantially similar.

The next section investigates the robustness of results to a non stationary specification of the PLM.

6.2 Robustness

The results sketched in Section 6.1 have been obtained by treating agents' PLM - the VAR(3) - as a stationary process. The learning mechanism proposed in the paper hinges on the capability of agents of testing whether their PLM supports the non-linear restrictions implied by the NKPC. The recent literature suggests that for certain type of restrictions tests in VAR models based on large sample approximations can exhibit distorted rejection rates (e.g. Dufour and Jouini, 2006), with the distortion increasing when the roots of the characteristic equation lay near the unit circle. This suggests, along with the diagnostics in Table 1 obtained over the initial period 1971:1-1983:4, that the robustness of the results obtained in Section 6.1 should be checked against non stationary specifications of the PLM.

Figure 3 plots the recursive estimates of the absolute value of the highest eigenvalues of the VAR(3) companion matrix over the period 1984:1-2005:4, providing *prima facie* descriptive evidence on how close the (inverse) roots of the characteristic equation are to the unit circle over the period used to evaluate the NKPC. More precisely, the estimated companion matrix has nine roots, four of which are complex conjugate. As emphasized by the graph, the absolute value of the two highest eigenvalues fluctuates in the range 0.96-0.99, whereas the absolute value of the third and forth eigenvalues fluctuates in the range 0.76-0.82. Overall, from Figure 3 it turns out that there might be room for the presence of two, possibly three unit roots in the system.

The suspect that a cointegrating relation (corresponding to two unit roots in the system) might characterize the VAR(3) can be ruled by looking at the graphs of Figure 4, where, after imposing the presence of a cointegrating relation (with a constant) in the system, we plotted the recursive (maximum likelihood) estimates of the cointegrating coefficients, normalizing the coefficient associated to inflation to one, over the period 1984:1-2005:4. The graph shows that it can be hardly concluded that e.g. euro area inflation and the wage share represent a stable cointegrating relation over the selected window.

If a VAR(3) with three stochastic trends represents a reasonable description of the level of the data until 2005, it makes sense to consider a DVAR(2) for $\Delta Z_t := (\Delta\pi_t \colon \Delta ws_t \colon \Delta i_t)'$ as agents' stationary PLM. Actually, as diagnostic tests computed on the DVAR(2) cast doubts on the uncorrelation of residuals, also a DVAR(3), corresponding to a VAR(4) for the levels, might be the candidate PLM not embodying stochastic trends. As detailed in Section 5, when π_t and ws_t are I(1) and not cointegrated, the empirical formulation of the forward-looking model of inflation dynamics which rules out I(1) trends is (31). For this reason, (31) is the NKPC tested with the DVAR systems.

The upper panels of Figure 5 plot the sequential LR tests of the CER imposed by the forward-looking model of inflation dynamics on the VAR(3) in levels (left panel) and its DVAR(2) counterpart (right panel), respectively; the lower panels report the same type of tests with reference to a VAR(4) and its DVAR(3) counterpart. The critical values of the sequential tests are taken from Table 1 in Inoue and Rossi (2005); note that the LR statistics and corresponding critical values in top-left panel of Figure 5 correspond to those already reported in Figure 2 and commented in Section 6.1.

Figure 5 shows that the NKPC is rejected from 1988 onward (77% of times) only when the DVAR(2) is considered; all the other specifications support the restrictions of the theory under the VAR-based learning dynamics. However, as the number of CER depends on the lag length, k, Figure 5 also reveals that the sequential LR test of CER might suffer of low power when relatively large values of k are considered, other than when the characteristic roots of the system are close to the unit circle. To sum up, the results obtained in Section 6.1 do not appear fully robust to specifications of agents' PLM which explicitly accounts for the presence of stochastic trends in the variables.

7 Concluding remarks

This paper provides a method for evaluating the NKPC under 'VAR expectations' and RLS learning. The approach is based on the idea that RE in the structural model are replaced by the forecasts stemming from a VAR including inflation, the wage share and possibly other variables. However, in this set-up the VAR serves not only as agents' PLM but also as the statistical platform upon which the validity of restrictions implied by the NKPC are evaluated recursively as the information set increases over time.

The proposed learning mechanism depends on the outcome of a likelihood-based test of the cross-equation restrictions that the NKPC imposes on the VAR parameters at any point in time. Agents learn under the null that the model of inflation dynamics is supported by the chosen VAR, or under the alternative. As new observations become available and the evaluation of the NKPC is repeated over time, agents may update the fraction of rejections and the fraction of non-rejections of the model, and check whether and in which periods their PLM is consistent with the structure of the forward-looking model of inflation dynamics.

The empirical analysis on euro area data points out that there is little room for the NKPC to serve as a reasonable description of inflation dynamics under the RE paradigm. On the contrary, if agents' PLM is approximated by a VAR estimated through RLS and the evaluation of the model is carried out over the period 1984-2005 with initial estimates obtained over the high inflation regime up to 1983, then a NKPC with the backward-looking parameter very close to zero seems to be consistent with agents' learning dynamics. This finding has remarkable implications on the optimal design of monetary policy. The empirical evidence, however, is less favorable to the NKPC when variables in the PLM are treated as non stationary and their stochastic trends are opportunely removed.

Appendix A

In this Appendix we prove that the CER (15) implied by the NKPC (1) amounts to a set of non-linear restrictions on of the form (16), which ensure the local identifiability of the system under the constraints. The result is reported in Proposition 1.

Proposition 1

Given the VAR (9) with companion matrix A, let $B_\pi' := g_\pi A$, $B_x' := g_x A$ and $B_a' := g_a A$ be the $1 \times n$ vectors, $n = pk$, containing the first-moment parameters associated with the marginal equation for π_t, x_t and a_t, respectively. Let $\tau = (\gamma_f \colon \gamma_b \colon \lambda)'$ be the 3×1 vector containing the structural parameters of the NKPC (1), and $\tau^0 := (\gamma_f^0 \colon \gamma_b^0 \colon \lambda^0)$, B_π^0, B_x^0 and B_a^0 their 'true' values. Let the scalar B_{x1}^0 be the first element of the vector B_x^0. Then, given the CER (15), if $\lambda^0 \neq 0$ and $B_{x1}^0 \neq (-\lambda^0/\gamma_f^0)$ when $\gamma_f^0 \neq 0$ and provided that $k \geq 2$ when $p \leq 3$ and $k \geq 1$ when $p \geq 4$, it exists an open set $\mathcal{D}$ in $\mathcal{R}^j$, $j := 2n+3$, containing $v^0 := (B_\pi^{0\prime} : B_a^{0\prime} : \tau^{0\prime})'$, and a unique (non-linear) differentiable function, f, $f : \mathcal{D} \longrightarrow \mathcal{R}^n$, such that

$$B_x^0 := f(B_\pi^0, B_a^0, \tau^0). \tag{36}$$

Remark

Proposition 1 ensures that the VAR is locally identifiable under the CER.

Proof.

First, note that the companion matrix of the VAR can be written as

$$A := \begin{bmatrix} B_\pi' \\ B_x' \\ B_a' \\ \vartheta' \end{bmatrix} \begin{array}{c} 1 \times n \\ 1 \times n \\ 1 \times n \\ (n-3) \times n \end{array}$$

where ϑ' is the lower block of the companion matrix, containing zeros and ones only, therefore

$$vec(A) = \begin{pmatrix} B_\pi \\ B_x \\ B_a \\ vec(\vartheta') \end{pmatrix} \begin{array}{c} n \times 1 \\ n \times 1 \\ n \times 1 \\ n(n-3) \times 1. \end{array} \tag{37}$$

Secondly, re-write the CER (15) in the form

$$B_\pi'(I_n - \gamma_f A) - \gamma_b g_\pi - \lambda B_x' = 0_{1\times n}$$

and apply the vec operator to both sides, obtaining

$$(I_n - \gamma_f A')B_\pi - \lambda B_x - \gamma_b g'_\pi = 0_{n\times 1}$$

which can be further compacted in the expression

$$H_{A,\tau} vec(A) - \gamma_b g'_\pi = 0_{n\times 1} \tag{38}$$

where the $n \times n^2$ matrix $H_{A,\tau}$ is defined as

$$H_{A,\tau} := [(I_n - \gamma_f A') \; : \; -\lambda I_n \; : \; 0_{n\times n} \; : \; 0_{n\times n(n-3)}] \; .$$

Observe that (38) formally defines a mapping of the type

$$g : \mathcal{S} \longrightarrow \mathcal{R}^n$$

where $\mathcal{S}$ is an open set in $\mathcal{R}^{n+j}$, such that

$$g(B_x : v) := 0_{n\times 1} \tag{39}$$

and $v := (B'_\pi : B'_a : \tau')'$; the mapping g in (39) summarizes the CER (15). Consider the $n \times n$ Jacobian

$$J(B_x : v) := \frac{\partial g(B_x : v)}{\partial B'_x}; \tag{40}$$

if $J(B_x : v)$ has full rank at the true point $(B_x^{0\prime} : v^{0\prime})'$, then by the implicit function theorem, there exists an open set $\mathcal{D}$ in $\mathcal{R}^j$ containing v^0, and a unique function $f : \mathcal{D} \longrightarrow \mathcal{R}^n$ with the following properties: (i) $f(v^0) =: B_x^0$; (ii) $g(f(v) : v) := 0$ for all v in D; (iii) f is continuosly differentiable in $\mathcal{D}$. Obviously, such a function corresponds to (36).

To prove that the function f exists, i.e. that the implict function theorem can be applied to the mapping g in (39), we must show that the Jacobian $J(B_x : v)$ in (40) is non-singular at $(B_x^{0\prime} : v^{0\prime})'$.

To simplify the computation of the Jacobian, express the first term on the right-hand-side of (38) as

$$H_{A,\tau} vec(A) \equiv \underset{n\times n^2}{H(B_x)} \; \underset{n^2\times 1}{d(B_x)}$$

where for notational convenience we have set $H_{A,\tau} \equiv H(B_x)$ and $d(B_x) \equiv vec[A(B_x)]$ to emphasize explicitly the dependence of $H_{A,\tau}$ and $vec(A)$ on the parameters in B_x.

It can be noticed that

$$\frac{\partial[H(B_x)d(B_x)]}{\partial B'_x} := \underset{n\times n^2}{H(B_x)} \underset{n^2\times n}{\frac{\partial d(B_x)}{\partial B'_x}} + \underset{n\times n^3}{\left[d(B_x)' \otimes I_n\right]} \underset{n^3\times n}{\frac{\partial vec[H(B_x)]}{\partial B'_x}}. \tag{41}$$

As regards the first addend of (41), given (37) one has

$$\frac{\partial d(B_x)}{\partial B_x'} = \frac{\partial vec(A)}{\partial B_x'} := \begin{bmatrix} \partial B_\pi / \partial B_x' \\ \partial B_x / \partial B_x' \\ \partial B_a / \partial B_x' \\ \partial vec(\vartheta') / \partial B_x' \end{bmatrix} = \begin{bmatrix} 0_{n\times n} \\ I_n \\ 0_{n\times n} \\ 0_{n(n-3)\times n} \end{bmatrix} := \underset{n^2\times n}{L}$$

so that

$$\begin{aligned} H(B_x)\ \frac{\partial d(B_x)}{\partial B_x'} \quad &: \quad = [(I_n - \gamma_f A') \ : \ -\lambda I_n \ : \ 0_{n\times n} \ : \ 0_{n\times n(n-3)}] \begin{bmatrix} 0_{n\times n} \\ I_n \\ 0_{n\times n} \\ 0_{n(n-3)\times n} \end{bmatrix} \\ &= \ -\lambda I_n. \end{aligned}$$

The derivative appearing in the second addend of (41) can be written as

$$\frac{\partial vec[H(B_x)]}{\partial B_x'} := K_{n^2,n} \frac{\partial vec[H(B_x)']}{\partial B_x'}$$

where $K_{n^2,n}$ is a $n^3 \times n^3$ commutation matrix (Magnus and Neudecker, 1999) such that $vec[H(B_x)] = K_{n^2,n} vec[H(B_x)']$. Thus

$$\begin{aligned} \frac{\partial vec[H(B_x)']}{\partial B_x'} &:= \begin{bmatrix} \partial vec(I_n - \gamma_f A)/\partial B_x' \\ -\partial vec(\lambda I_n)/\partial B_x' \\ 0_{n^2\times n} \\ 0_{n^2(n-3)\times n} \end{bmatrix} = \begin{bmatrix} -\gamma_f \partial vec(A)/\partial B_x' \\ 0_{n^2\times n} \\ 0_{n^2\times n} \\ 0_{n^2(n-3)\times n} \end{bmatrix} \\ &= \begin{bmatrix} -\gamma_f \frac{\partial d(B_x)}{\partial B_x'} \\ 0_{n^2\times n} \\ 0_{n^2\times n} \\ 0_{n^2(n-3)\times 3} \end{bmatrix} := \begin{bmatrix} -\gamma_f L \\ 0_{n^2\times n} \\ 0_{n^2\times n} \\ 0_{n^2(n-3)\times 3} \end{bmatrix}, \end{aligned}$$

hence

$$\underset{n\times n^3}{\left[d(B_x)' \otimes I_n\right]} \ \underset{n^3\times n}{\frac{\partial vec[H(B_x)]}{\partial B_x'}} := \underset{n\times n^3}{\left[d(B_x)' \otimes I_n\right]} \ \underset{n^3\times n^3}{K_{n^2,n}} \ \underset{n^3\times n}{\begin{bmatrix} -\gamma_f L \\ 0_{n^2\times n} \\ 0_{n^2\times n} \\ 0_{n(n-3)\times n} \end{bmatrix}}$$

$$= \left[I_n \otimes d(B_x)'\right] \begin{bmatrix} -\gamma_f L \\ 0_{n^2\times n} \\ 0_{n^2\times n} \\ 0_{n(n-3)\times 3} \end{bmatrix} = \begin{bmatrix} d(B_x)' & 0_{1\times n^2} & \cdots & 0_{1\times n^2} \\ 0_{1\times n^2} & d(B_x)' & \cdots & 0_{1\times n^2} \\ \vdots & & \ddots & \\ 0_{1\times n^2} & 0_{1\times n^2} & \cdots & d(B_x)' \end{bmatrix} \begin{bmatrix} -\gamma_f L \\ 0_{n^2\times n} \\ 0_{n^2\times n} \\ 0_{n(n-3)\times 3} \end{bmatrix}$$

$$= -\gamma_f \underset{n\times n}{\begin{bmatrix} d(B_x)'L \\ 0_{1\times n} \\ \vdots \\ 0_{1\times n} \end{bmatrix}} \tag{42}$$

where we have used the following properties of the commuation matrix: (i) $[d' \otimes I_n]\ K_{n^2,n} = K_{1,n}\,[I_n \otimes d']$; (ii) $K_{1,n} = I_n$, see Magnues and Neudecker (1999), pp 46-47. By focusing on the first row of the matrix (42), it turns out that using (37), the definition of $d(B_x)$ and the definition of the matrix L, one has

$$\begin{aligned} d(B_x)'L &\equiv vec(A)'L \equiv (B_\pi' : B_x' : B_a' : vec(\vartheta')') \begin{bmatrix} 0_{n\times n} \\ I_n \\ 0_{n\times n} \\ 0_{n(n-3)\times n} \end{bmatrix} \\ &= B_x' := (B_{x1} : B_{x2} : ... : B_{xn}), \end{aligned}$$

where each B_{xh}, $h := 1, ..., n$ is the hth element of B_x. We have thus proved that the matrix in (42) has the structure:

$$-\gamma_f \begin{bmatrix} d(B_x)'L \\ 0_{1\times n} \\ \vdots \\ 0_{1\times n} \end{bmatrix} := -\gamma_f \begin{bmatrix} B_{x1} & B_{x2} & \cdots & B_{xn} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

To sum up, the Jacobian (40) is given by

$$J(B_x : v) := -\lambda \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} - \gamma_f \begin{bmatrix} B_{x1} & B_{x2} & \cdots & B_{xn} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$= -1 \begin{bmatrix} \lambda + \gamma_f B_{x1} & \gamma_f B_{x2} & \cdots & \gamma_f B_{xn} \\ 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda \end{bmatrix}.$$

By evaluating $J(B_x : v)$ at the true parameter values, yields:

$$J(B_x^0 : v^0) := -1 \begin{bmatrix} \lambda^0 + \gamma_f^0 B_{x1}^0 & \gamma_f^0 B_{x2}^0 & \cdots & \gamma_f^0 B_{xn}^0 \\ 0 & \lambda^0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda^0 \end{bmatrix}.$$

Hence $J(B_x^0 : v^0)$ is non-singular if $\lambda^0 \neq 0$ and $B_{x1}^0 \neq (-\lambda^0/\gamma_f^0)$ when $\gamma_f^0 \neq 0$. Observe that the VAR (9) subject to (36) has $2n+$ 3 free parameters, hence the total number of CER is $3n - (2n+3) = n - 3 = pk - 3$. This implies that $k \geq 2$ when $p \leq 3$ and $k \geq 1$ when $p \geq 4$ for the restrictions to be binding. This completes the proof. ■

References

Baillie, R. T. (1989), Econometric tests of rationality and market efficiency, *Econometric Reviews* 8, 151-186.

Batini, N. (2006), Euro are inflation persistence, *Empirical Economic* 31, 977 - 1002.

Benigno, P., López-Salido, J.D. (2006), Inflation persistence and optimal monetary policy in the euro area, *Journal of Money Credit and Banking* 38, 587-614.

Branch, W.A. (2004), The theory of rationally heterogeneous expectations: evidence from survey data on inflation expectations, *Economic Journal* 114, 592-621.

Branch, W.A., Evans, G.W. (2006), A simple recursive forecasting model, *Economic Letters* 91, 158-166.

Brayton, F., Eileen, M., Reifschneider, Tinsley, P., Williams, J. (1997), The role of expectations in the FRB/US macroeconomic model, *Federal Reserve Bulletin* 83 (April), 227-245.

Bullard, J., Mitra, K. (2002), Learning about monetary policy rules, *Journal of Monetary Economics* 49, 1105-1129.

Campbell, J. Y., Shiller, R. J. (1987), Cointegration and tests of present value models, *Journal of Political Economy* 95, 1062–1088.

Carceles-Poveda, E., Giannitsarou (2006), Adaptive learning in practise, *Journal of Economic Dynamics and Control,* forthcoming.

Del Negro, M., Schorfheide, F. (2006), Monetary policy analysis with potentially misspecified models, paper presented at the 17th EC^2 Meeting, "*The Econometrics of Monetary Policy and Financial Decision-Making*", December 15-16, 2006, Rotterdam.

Doornik, J. A, Hendry, D. F. (2001), *Modelling Dynamic Systems Using PcGive*, Timberlake Consultants Press, London.

Dufour, J.-M, Jouini, T. (2006), Finite-sample simulation-based inference in VAR models with application to Granger causality testing, *Journal of Econometrics* 135, 229-254.

Evans, G. W. and Honkapohja, S. (1999), *Learning dynamics*, in Handbook of Macroeconomics 1A, Chap. 7

Evans, G. W., Honkapohja, S. (2001), *Learning and expectations in macroeconomics*, Princeton University Press.

Evans, G. W., Honkapohja, S. (2003*a*), Adaptive learning and monetary policy design, *Journal of Money, Credit and Banking* 35, 1045-1072.

Evans, G. W., Honkapohja, S. (2003*b*), Expectations and the stability problem for optimal monetary policies, *Review of Economic Studies* 70, 807-824

Fagan, G., Henry, G. and Mestre, R. (2001), An area-wide model (awm) for the Euro area, *European Central Bank, Working Paper* No. 42.

Fanelli, L. (2006*a*), Dynamic adjustment cost models with forward-looking behavior, *Econometrics Journal* 9, 23-47.

Fanelli, L. (2006*b*), Multi-equational linear quadratic adjustment cost models with rational expectations and cointegration, *Journal of Economic Dynamics and Control* 30, 445-456.

Fanelli, L. (2006*c*), Testing the New Keynesian Phillips Curve through Vector Autoregressive models: Results from the Euro area, *Working Paper*, University of Bologna.

Fuhrer, J. (1997), The (un)importance of forward-looking behavior in price specifications, *Journal of Money Credit and Banking* 29, 338-350.

Fuhrer, J., Moore, G. (1995), Inflation persistence, *Quarterly Journal of Economics* 110, 127-159.

Galí, J., Gertler, M. (1999), Inflation dynamics: a structural econometric analysis, *Journal of Monetary Economics* 44, 195-222.

Galí, J., Gertler M. and Lopez-Salido, J.D. (2001), European inflation dynamics, *European Economic Review* 45, 1237-1270.

Goffe, W. L., Ferrier, G. D. and Rogers, J. (1994), Global optimization of statistical functions with simulated annealing, *Journal of Econometrics* 60, 65-99.

Hansen, L.P., Sargent, T.J. (1980), Formulating and estimating dynamic linear rational expectations models, *Journal of Economic Dynamics and Control* 2, 7-46.

Hansen, L.P., Sargent, T.J. (1981), *Linear rational expectations models for dynamically interrelated variables*, in: Lucas R.E.Jr., Sargent, T.J. (dds.), *Rational Expectations and Econometric Practise*. University of Minnesota Press, Minneapolis, pp. 127-156.

Inoue, A., Rossi, B. (2005), Recursive predictability tests for real-time data, *Journal of Business and Economic Statistics* 23, 336-345.

Ireland, P. (2004), A Method for taking models to the data, *Journal of Economic Dynamics and Control* 28, 1205-1226.

Johansen, S. (1996), *Likelihood-based inference in cointegrated Vector Auto-Regressive models*, Oxford University Press, Second revised version, Oxford.

Johansen, S. (2006), Confronting the economic model with the data, in Colander, D. (ed): *Post Walrasian Macroeconomics*, Cambridge University Press, Cambridge, pp. 287-300.

Johansen, S., Swensen, A. R. (1999), Testing exact rational expectations in cointegrated vector autoregressive models, *Journal of Econometrics* 93, 73-91.

Kozicki, S., Tinsley, P. A. (1999), Vector rational error correction, *Journal of Economic Dynamics and Control* 23, 1299-1327.

Kurmann, A. (2006), Maximum likelihood estimation of dynamic stochastic theories with an application to New Keynesian pricing, *Journal of Economic Dynamics and Control*, forthcoming.

Lindé, J. (2005), Estimating New Keynesian Phillips curves: A full information maximum likelihood approach, *Journal of Monetary Economics* 52, 1135-1149.

Lubik, T., Schorfheide, F. (2004), Testing for indeterminacy: an application to U.S. monetary policy, *American Economic Review* 2004, 94, 190-217.

Marcet, A., Sargent, T.J. (1992), *The convergence of Vector Autoregressions to rational expectations equilibria*, in A. Vercelli and N. Dimitri (eds.), *Macroeconomics: A Survey Of Reserach Strategies*, Oxford, Oxford University Press 139-164.

Mavroeidis, S. (2004), Weak identification of forward-looking models in monetary economics, *Oxford Bulletin of Economics and Statistics* 66 (Supplement), 609-635.

Mavroeidis, S. (2005), Identification issues in forward-looking models estimated by GMM, with an application to the Phillips curve, *Journal of Money Credit and Banking* 37, 421-448.

Mavroeidis, S. (2006), Testing the New Keynesian Phillips Curve without assuming identification, paper presented at the 17th EC^2 Meeting, "*The Econometrics of Monetary Policy and Financial Decision-Making*", December 15-16, 2006, Rotterdam.

McCallum, B. T. (1983), On non-uniqueness in rational expectations models: an attempt at perspective, *Journal of Monetary Economics* 11, 139-168.

McCallum, B. T. (2003), The unique minimum state variable RE solutuion is E-stable in all well formulated models, *NBER, Working Paper* No. 9960.

Milani, F. (2005*a*), Adaptive learning and inflation persistence, *Working Paper*, University of California, Irvine.

Milani, F. (2005*b*), Expectaions, learning and macroeconomic persistence, *Working Paper*, University of California, Irvine.

Muth, J.F. (1961), Rational expectations and the theory of price movements, *Econometrica* 29, 315-335.

Nason, J. M., Smith, G. W. (2005), Identifying the New Keynesian Phillips curve, *Research Department, Federal Reserve Bank of Atlanta.*

Orphanides, A., Williams, J.C. (2005), The decline of activist stabilization policy: Natural rate mispercpections, learning, and expectations, *Journal of Economic Dynamics and Control* 29, 1927-1950.

O'Relly, G., Whelan, K. (2005), Has euro-area inflation persistence changed over time ?, *Review of Economics and Statistics* 87, 709-720.

Pesaran, H. M. (1987), *The limits to rational expectations*, Basil Blackwell, Oxford.

Primicieri, G. (2006), Why inflation rose and fell: policy-makers' beliefs and U.S. postwar stabilization policy, *Quarterly Journal of Economics*, March 867-901.

Roberts, J. M. (1995), New Keynesian economics and the Phillips curve, *Journal of Money, Credit, and Banking* 27 (4), 975-984.

Roberts, J. M. (1997), Is inflation sticky?, *Journal of Monetary Economics* 39, 173-196.

Ruud, J., Whelan, K. (2005*a*), Does labor's share drive inflation ?, *Journal of Money Credit and Banking* 37, 297-312.

Ruud, J., Whelan, K. (2005*b*), New tests of the New Keynesian Phillips curve, *Journal of Monetary Economics* 52, 1167-1181.

Ruud, J., Whelan, K. (2006), Can rational expectations sticky-price models explain inflation dynamics?, *American Economic Review* 96, 303-320.

Sargent, T.J. (1979). A note on the maximum likelihood estimation of the rational expectations model of the term structure, *Journal of Monetary Economics* 5, 133-143.

Sargent, T.J. (1987), *Macroeconomic theory* (second edition), San Diego, Academic Press.

Sargent, T.J. (1999), *The conquest of American inflation*, Princeton University Press, Princeton.

Sbordone, A.M. (2002), Prices and unit labor costs: a new test of price strickiness, *Journal of Monetary Economics* 49, 265-292.

Sbordone, A.M. (2005), Do expected future marginal costs drive inflation dynamics ?, *Journal of Monetary Economics* 52, 1183-1197.

Smets, F., Wouters, R. (2003), An estimated dynamic stochastic general equilibrium model of the Euro area, *Journal of the European Economic Association* 1(5), 1123-1175.

Woodford, M. (2003), *Interest and prices: Foundations of a theory of monetary policy*, Princeton Universiyt Press, Princeton.

Figures

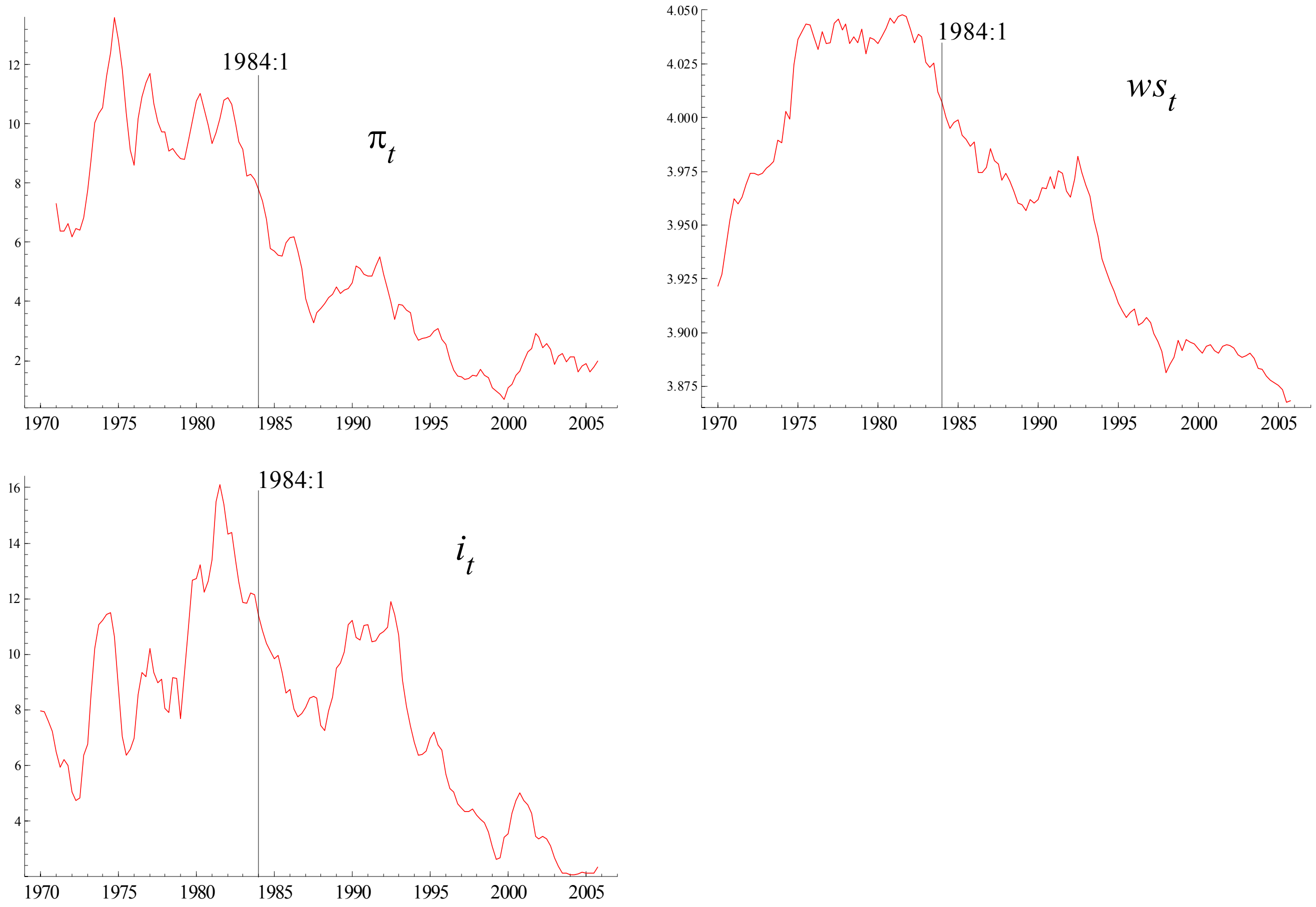

Figure 1: Euro area inflation (π_t), wage share (ws_t) and short term nominal interest rate (i_t). Quarterly data 1971:1-2005:4. The vertical line at 1984:1 divides the sample used to produce initial estimates of the VAR from that used to evaluate the NKPC.

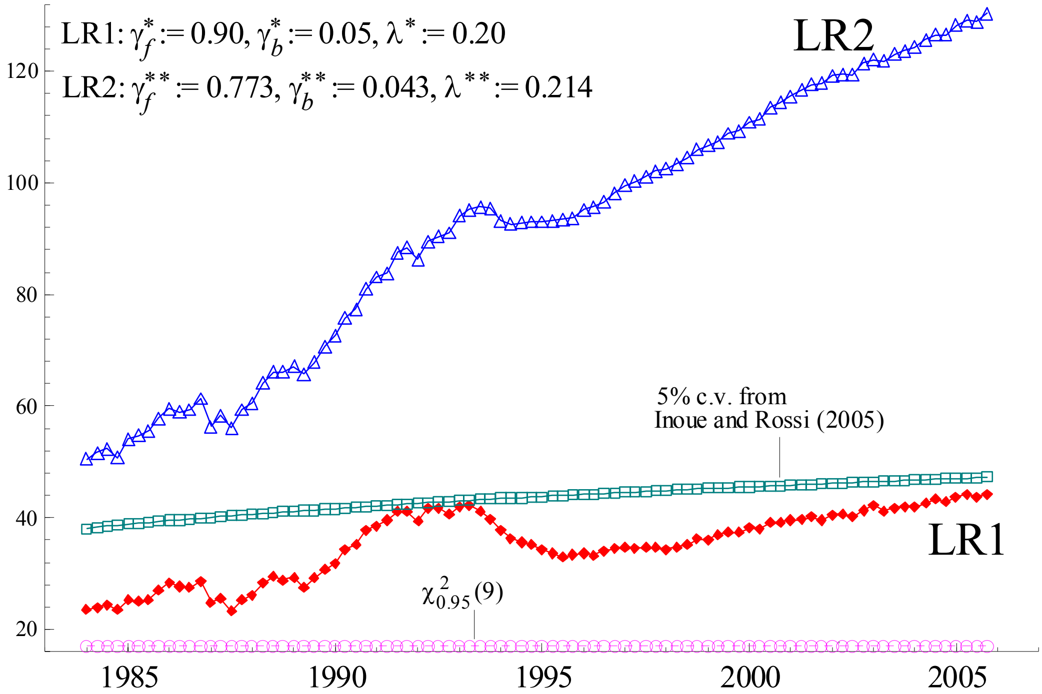

Figure 2: Sequence of LR tests of the CER implied by the NKPC on the VAR(3). LR1 (⧫) is obtained by maximing the log-likelihood of the VAR(3) under the CER with $(\gamma_f : \gamma_b : \lambda)'$ fixed as described in the top of the graph. LR2 ($\triangle$) is obtained by maximing the log-likelihood of the VAR(3) under the CER with $(\gamma_f : \gamma_b : \lambda)'$ fixed as in Galì et al. (2001). $\square$ = 5% critical value taken from Inoue and Rossi (2005); $\circ = \chi^2_{0.95}(9)$ is the 0.95 quantile of a χ^2 with 9 d.f.

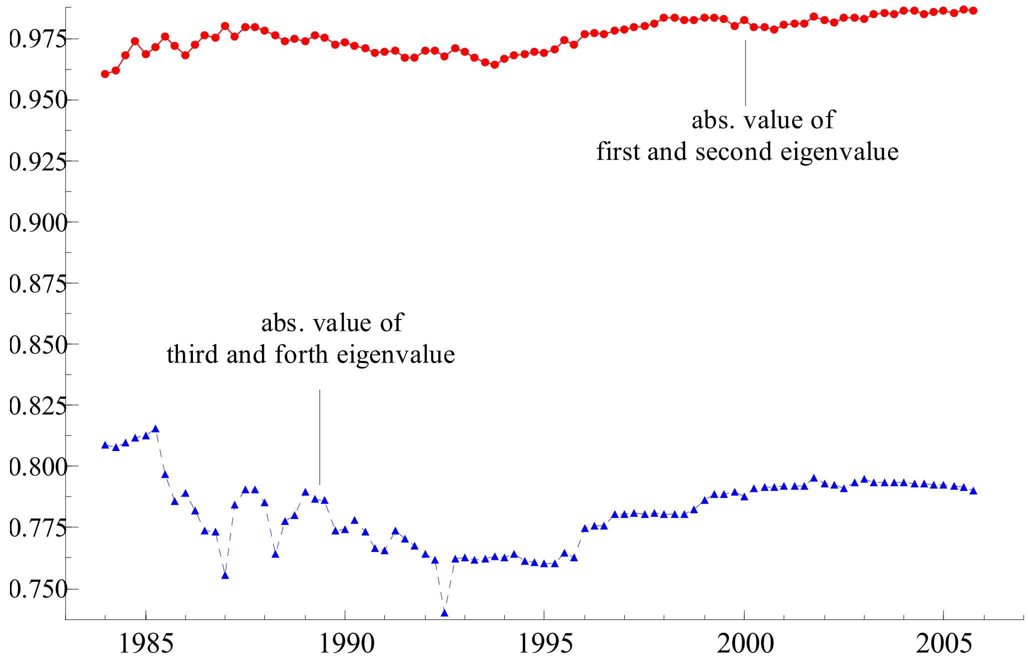

Figure 3: recursive estimates over the period 1984:1-2005:4 of the absolute value of the highest eigenvalues of the VAR(3) companion matrix.

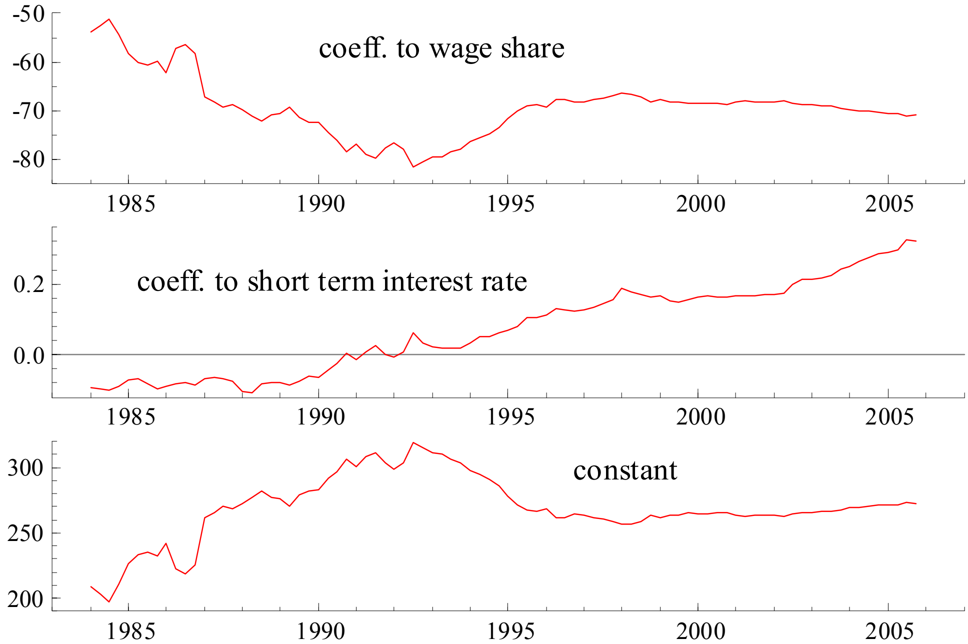

Figure 4: recursive estimates (maximum likelihood) over the period 1984:1-2005:4 of the coefficients of the cointegration relation between π_t (normalized to one), ws_t, i_t and a constant obtained by imposing rank cointegration rank to one in the estimated VAR(3).

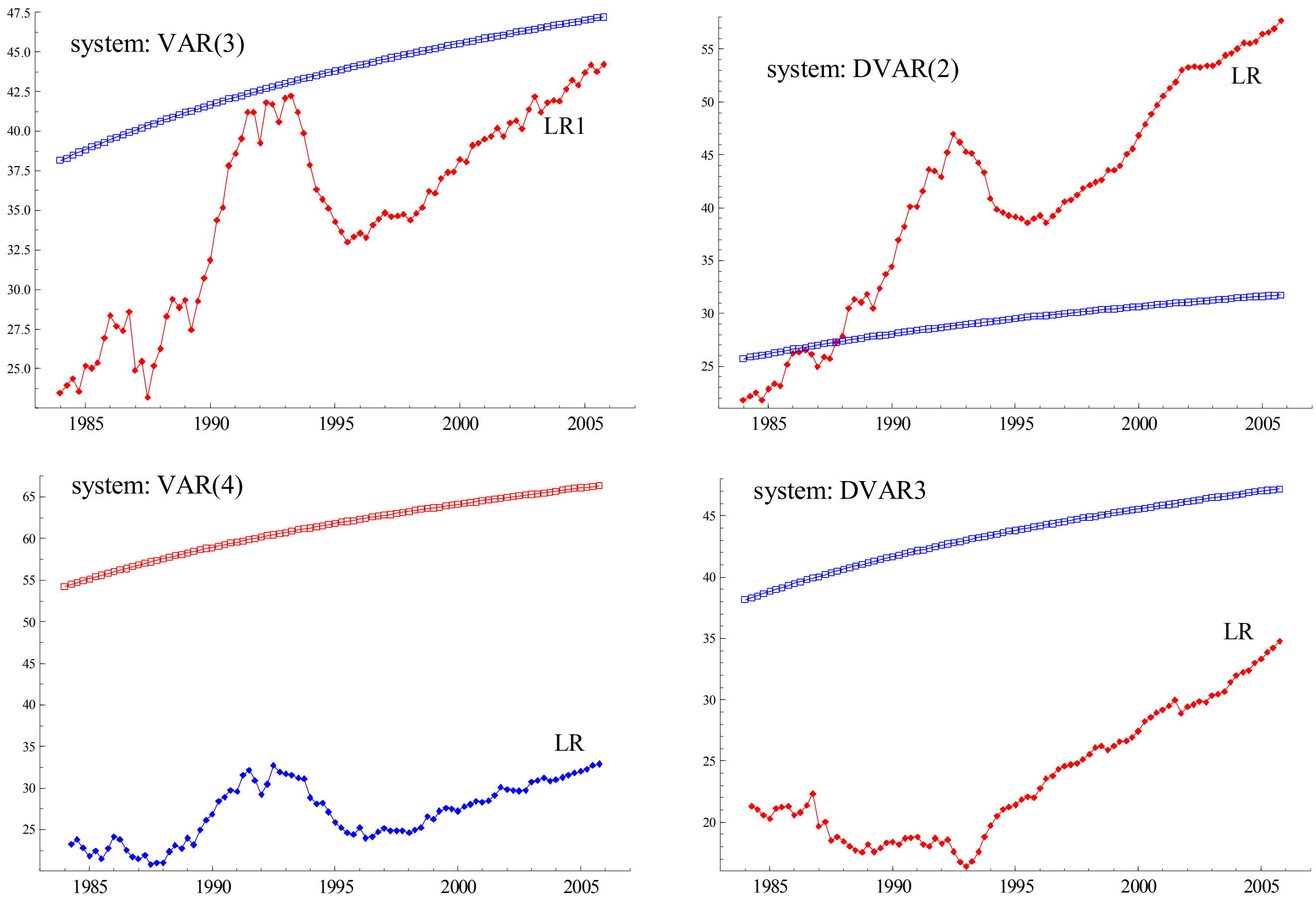

Figure 5: Sequence of LR tests (♦) of the CER implied by the NKPC with respect to different specifications of the PLM in levels (left panels) and in first differences (right panels). □ = 5% critical values taken from Inoue and Rossi (2005).

A Tables

VAR(3): $Z_t = (\pi_t : ws_t : i_t)'$ 1971:1- 1983:4
Vector AR 1-3 test: $F(36,74) := \underset{[0.37]}{1.08}$
Vector normality test: $\chi^2(6) := \underset{[0.24]}{8.02}$
Highest eigenvalues of companion matrix: $0.95 \pm 0.073i$

Cointegration rank test		
$H_0 : r \leq j$	Trace	p-val
j=0	35.97	0.039
j=1	13.73	0.31
j=2	5.16	0.28

Table 1: Upper panel: vector test for residual autocorrelation and normality, see Doornik and Hendry (2001). Middle panel: highest estimated eigenvalues of the VAR companion matrix. Lower panel: Johansen's (1996) LR trace test for cointegration rank with corresponding p-values.